Routledge
Taylor & Francis Group

Check for updates

Reliability of the French-Canadian adaptation of the Personality Assessment Inventory: Medical–legal implications

Eliyas Jeffay[a], Angela Sekely[a], Michel Lacerte[b] and Konstantine K. Zakzanis[a]

[a]*Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada;*
[b]*Department of Physical Medicine and Rehabilitation, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada*

The Personality Assessment Inventory (PAI) is a reliable multidimensional psychometric inventory that is increasingly being used in the medical–legal context. To date, 18 language adaptations of the PAI exist, yet only the Spanish, Greek and German language versions have been examined psychometrically. This study evaluated the psychometric properties of the French-Canadian version of the PAI by comparing mean scale and subscale scores between the French-Canadian and English language versions, and analyzing the internal consistency and mean item inter-correlations (MICs) of each version in a sample of 50 bilingual university students. Cronbach's alphas ranged from $-.57$ to $.80$ in the French-Canadian version and from -1.10 to $.83$ in the English version, with most scales being below $.70$, indicating inadequate internal consistency. In addition, most of the MICs were below $.20$, indicating a lack of item homogeneity. Caution is given to this adaptation of the PAI in the medical–legal context.

Key words: bilingual; language adaptation; medical–legal; Personality Assessment Inventory; psychological assessment; psychometrics; reliability.

The generalizability of psychological tests with specific populations is an insidious problem in clinical psychology and may have significant implications in the context of a medical–legal examination. For example, if psychologists are asked to objectively substantiate the breadth, severity and veracity of subjective symptomatology and come to a diagnostic opinion, it is essential that such opinions are based on firm scientific grounds in order to meet legal standards and be accepted by the courts. This is especially important since psychologists are asked to suggest/comment on the efficacy of treatment, determine disability benefit or comment on the permanence or seriousness of a psychological injury, all in the ultimate context of assisting the trier of fact in a medical–legal setting. The *Standards of Educational and Psychological Testing* state that translating a measure into another language does not ensure the construct measured remains comparable to the original test (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014). As such, the examination of language adaptations is an essential part of the study of cultural differences and similarities (Ellis, 1989). This does not, of course, discount the complex and distinct sub-cultural groups who speak the same language (e.g. French speakers who are

Correspondence: Eliyas Jeffay, Department of Psychology, University of Toronto, Scarborough, 1265 Military Trail, Scarborough, ON, M1C 1A4, Canada. Email: eliyas.jeffay@utoronto.ca

Moroccan, Congolese, Belgian, etc.). Cheung (2009) explained that personality instruments developed in Western cultures are often generalized to other cultural groups with the faulty assumption that these measures are valid for all groups. When a measure is adapted for a population that differs qualitatively from the one for which it was originally developed, the reliability and validity of the test must be evaluated before it can be clinically utilized (Butcher, Derksen, Sloore, & Sirigatti, 2003; Candell & Hulin, 1986; Cheung, 2009; Geisinger, 1994; Sireci & Berberoglu, 2000) and, hence, employed in the context of a medical–legal examination. Despite these recommendations, research on the language adaptations of the Personality Assessment Inventory (PAI; Morey, 1991, 2007) has been limited.

The PAI is a self-report instrument that yields a broad range of clinically relevant information, and is a widely utilized test measure of personality and psychopathology in medical–legal examinations. It was developed using a rational and quantitative method of scale development. The rational criterion emphasizes theoretically informed choices when developing items, as opposed to empirically based instruments such as the Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). The PAI consists of 344 items that constitute four sets of non-overlapping scales: (a) four validity scales: Inconsistency, Infrequency, Negative Impression Management and Positive Impression Management; (b) 11 clinical scales: Somatic Complaints, Anxiety, Anxiety-Related Disorders, Depression, Mania, Paranoia, Schizophrenia, Borderline Features, Antisocial Features, Alcohol Problems and Drug Problems; (c) five treatment scales: Aggression, Suicidal Ideation, Stress, Nonsupport and Treatment Rejection; and (d) two interpersonal scales: Dominance and Warmth. Several advantages of the PAI include its brevity, lower reading level

requirements, focus on diagnostic concepts and attention to clinical management issues. The acronyms for the scales and subscales are presented in Table 1.

In addition to the original English version created in the US, the PAI is available in the following 18 languages: Arabic, Brazilian Portuguese, Bulgarian, Chinese, Filipino, French Canadian, German, Greek, Icelandic, Korean, Norwegian, Polish, Serbian, Slovene, Spanish, Swedish, Turkish and Vietnamese. However, Cheung and colleagues (1996) cautioned against the direct interpretation from the original normative data, because the culturally different examinee may be misjudged and responses deemed invalid. Despite these warnings, only the Spanish, German and Greek languages have been evaluated empirically. All of these studies examined internal consistency using Cronbach's alphas (α), in that the PAI items from the same scale are assumed to measure the same construct. In the original English version, Morey (2007) reported moderate αs on the validity scales (normative sample = .45 to .71; college students = .22 to .73; and clinical patients = .23 to .77), with Inconsistency (INC) and Infrequency (INF) tending to be lower than other scales. In addition, internal consistency estimates were consistently high for the clinical (normative = .74 to .90; college students = .66 to .89; and clinical patients = .82 to .93), treatment (normative = .72 to .85; college students = .69 to .89; and clinical patients = .79 to .90), and interpersonal (normative = .78 and .79; college students = .80 and .81; and clinical patients = .82 and .83) scales. The Spanish version, validated on bilingual Mexican Americans (Rogers, Flores, Ustad, & Sewell, 1995), revealed low αs for the validity scales (.29 to .70) and modest αs for the clinical (.40 to .82), treatment (.40 to .82) and interpersonal (.41 and .71) scales. Another Spanish version, validated on Argentineans (Stover, Solano, & Liporace, 2015), revealed modest reliability coefficients for the validity scales (.52 to .70), although INC and INF were not reported.

Table 1. Acronyms for PAI scales and subscales.

Scale acronym	Scale name	Subscale acronym	Subscale name
Validity			
INC	Inconsistency		
INF	Infrequency		
NIM	Negative Impression		
PIM	Positive Impression		
Clinical			
SOM	Somatic Complaints		
		SOM-C	Conversion
		SOM-S	Somatization
		SOM-H	Health Concerns
ANX	Anxiety		
		ANX-C	Cognitive
		ANX-A	Affective
		ANX-P	Physiological
ARD	Anxiety-Related Disorders		
		ARD-O	Obsessive-Compulsive
		ARD-P	Phobias
		ARD-T	Traumatic Stress
DEP	Depression		
		DEP-C	Cognitive
		DEP-A	Affective
		DEP-P	Physiological
MAN	Mania		
		MAN-A	Activity Level
		MAN-G	Grandiosity
		MAN-I	Irritability
PAR	Paranoia		
		PAR-H	Hypervigilance
		PAR-P	Persecution
		PAR-R	Resentment
SCZ	Schizophrenia		
		SCZ-P	Psychotic Experiences
		SCZ-S	Social Detachment
		SCZ-T	Thought Disorder
BOR	Borderline Features		
		BOR-A	Affective Instability
		BOR-I	Identity Problems
		BOR-N	Negative Relationships
		BOR-S	Self-Harm
ANT	Antisocial Features		
		ANT-A	Antisocial Behaviors
		ANT-E	Egocentricity
		ANT-S	Stimulus Seeking
ALC	Alcohol Problems		
DRG	Drug Problems		

(*Continued*)

Table 1.   (*Continued*).

Scale acronym	Scale name	Subscale acronym	Subscale name
Treatment			
AGG	Aggression		
		AGG-A	Aggressive Attitude
		AGG-V	Verbal Aggression
		AGG-P	Physical Aggression
SUI	Suicidal Ideation		
STR	Stress		
NON	Nonsupport		
RXR	Treatment Rejection		
Interpersonal			
DOM	Dominance		
WRM	Warmth		

Note: PAI = Personality Assessment Inventory.

In addition, high αs were reported for the clinical scales (.70 to .86), and modest αs were reported for the treatment (.60 to .82) and interpersonal (.68 and .71) scales. The German version (Groves & Engel, 2007) had similar reliability coefficients, with validity scales ranging from .26 to .73, clinical scales ranging from .63 to .91, treatment scales ranging from .70 to .87, and interpersonal scales being .72 and .76. The Greek version (Lyrakos, 2011) had high αs on the validity scales (healthy = .85 to .86; inpatients and outpatients = .86 to .94; and outpatients = .97 to .99), although values for INC and INF were not reported. In addition, high αs were reported on the clinical (healthy = .74 to .95; inpatient and outpatients = .78 to .96; and outpatients = .76 to .99), treatment (healthy = .74 to .92; inpatient and outpatients = .86 to .99; and outpatients = .80 to .89), and interpersonal (healthy = .83 and .83; inpatient and outpatients = .76 and .85; and outpatients .86 and .99) scales. To summarize, the literature suggests that although the reliability coefficients for the INC and INF validity scales appear to be low, all other scales appear to be of adequate psychometric property. However, additional research is warranted to support the clinical and medical–legal utility of the PAI for the remaining 15 languages.

Whereas many instruments of high clinical utility have been created in the US, relatively few measures have been effectively translated and adapted for use in the Canadian population (Jeanrie & Bertrand, 1999). This is especially important when one considers French-Canadian respondents. In French-Canadian samples, not only is there a potential for language differences, but there is also an added complexity of cultural differences. These factors act as external sources of variance, making it less likely that true scores are estimated by the testing instrument. As Jeanrie and Bertrand (1999) explained, 'while language and potential cultural differences might already affect scores when one compares French-Canadians to American norms, careless translations can introduce additional biases that will decrease the validity of one's test scores' (p. 278). While translated tests are commercially available for this population, most translated tests are provided without any information pertaining to how the test has been translated or validated. Notably, though a French-Canadian version of the PAI exists, it has not been validated in any group of French-speaking Canadians.

Thus, it is essential that the French-Canadian version of the PAI be validated. To test for construct equivalence (i.e. the

generalizability of the test to other cultures) in adapted personality measures, the use of bilingual test–retest studies has been recommended (Butcher, Mosch, Tsai, & Nezami, 2006; Sireci & Berberoglu, 2000). In this design, a group of bilingual individuals in the target culture (i.e. French Canadians) would take both the original form of the test and the translated version of the test. These two tests are then compared to determine whether the scales are operating in the same manner in both language versions (see Butcher et al., 2003; Chen & Bond, 2010). Although it is highly unlikely that bilingual test-takers are equally proficient in two languages, having individuals who are literate in both languages complete both versions of the test has several advantages. One advantage is that the same examinees are responding to both language versions, simultaneously accounting for individual differences, group proficiency differences and item translation differences. Thus, any differences between language versions can be attributed to translational rather than cultural differences. For a detailed review on the use of bilingual respondents to evaluate translated tests, see Sireci and Berberoglu (2000). In addition, it reflects the real-world literacy of bilingual test-takers in clinical settings.

Accordingly, we sought to examine the psychometric properties of the French-Canadian PAI scales and subscales in a bilingual sample. The bilingual test–retest study design was used to evaluate whether the scales function in a similar way across translated adaptations.

Method
Participants

A total of 56 university student participants were recruited from the University of Toronto Scarborough Campus as part of their academic fulfilment in an introduction to psychology course. Participants received partial course credit as compensation for their involvement in the study. The participants were primarily female (75.00%), with 19.60% male, and 5.40% did not wish to disclose gender. Age of the participants ranged between 17 and 36 years ($M = 19.50$, $SD = 3.09$). The majority of participants were in their first year of university (69.60%). In addition, 17.90% were in second year, 1.80% were in third year, 8.90% were in fourth year, and 1.80% were in sixth year. According to a 10-point scale (see next section), participants were both fluent ($M = 9.55$. $SD = 0.81$) and literate ($M = 9.64$, $SD = 0.75$) in English, and fluent ($M = 8.54$, $SD = 1.26$) and literate ($M = 8.91$, $SD = 1.00$) in French. Ethics approval was granted by way of the University of Toronto research ethics board. Participants were treated with the Tri-Council Policy Statement (second edition) in that informed consent was obtained, and the participant was entirely aware that they could withdraw their participation in the study at any given time without consequence (Canadian Institutes of Health Research, Natural Sciences & Engineering Research Council of Canada, & & Social Sciences & Humanities Research Council of Canada, 2010).

Exclusions

Exclusion criteria consisted of (a) a positive history of illicit drug use, known neurological disease, psychiatric illness and/or past head injury; (b) less than a Grade 4 reading ability on the Wide Range Achievement Test–Fourth Edition Reading subtest (WRAT–4–R; Wilkinson & Robertson, 2006); and/or (c) a rating of less than 7/10 on a self-reported English fluency/literacy, French fluency/literacy items on the Language Background and Use Questionnaire. These inclusion criteria were selected because they can affect test performance. A total of two participants were excluded from analysis due to the criterion (a). In addition, four participants were excluded from analyses because they endorsed less than 7/10 on one of the items on the Language Background and Use Questionnaire (see Table 2). No individuals were excluded from the sample according to the inclusion criterion (b).

Table 2. Language fluency and literacy questionnaire self-ratings.

Response	Fluency		Literacy	
	Frequency	%	Frequency	%
English				
6/10	0	0.00	0	0.00
7/10	1	1.80	1	1.80
8/10	8	14.30	6	10.70
9/10	6	10.70	5	8.90
10/10	41	73.20	44	78.60
French				
5/10	1	1.80	0	0.00
6/10	2	3.60	1	1.80
7/10	9	16.10	2	3.60
8/10	14	25.00	18	32.10
9/10	14	25.00	15	26.80
10/10	16	28.60	20	35.70

Thus, a total of 50 participants were included in the current study (Table 2).

Measures

The PAI is a self-report multiscale instrument containing 344 items organized into four validity scales, 11 clinical scales, five treatment scales and two interpersonal scales (described earlier). Raw scores were entered into the PAI Software Portfolio (2000), where they were converted into t scores ($M = 50$, $SD = 10$) based on comparison to a normative reference sample. As per the manual, a t score that is greater than 50 indicates that the individual has endorsed items that reflect a specific construct to a greater degree than what is typical in the general population. Moreover, a t score that is greater than 70 represents a degree of symptoms that is unusual in the general population and most likely indicates clinical significance. The translated version of the PAI was purchased from Psychological Assessment Resources (PAR) Inc. and used with permission for the purpose of this study. As per PAR, the English version of the PAI was translated into French by one team. The French translation was then back-translated to English by a bilingual individual unfamiliar with the English version of the test. This back-translation was then forwarded to the test publishing company for review and approval. No translation problems were reported by PAR with any of the items. Both versions of the test were used with permission by PAR for the purposes of this study.

Procedure

Following consent to participate, demographic data were collected by way of a brief interview form. This was given to ascertain demographic variables, such as age, gender, years of education, first language, second language and handedness, and to screen for any psychiatric, neurological or relevant pathology. Following this, the Language Background and Use Questionnaire, a brief self-report measure assessing the participant's language abilities, was administered to determine the individual's level of competency in their first and second language. Next, the WRAT–4–R was administered to determine whether the participants had sufficient education/reading level (minimum score of 45; Grade 4) for completing the PAI, as per the manual (Morey, 2007). The participants were then administered the PAI in either English (Morey, 1991, 2007) or French.

Participants were then administered the version they did not complete in the first administration (i.e. either English or French). The order of administration was randomized to dispel any order and/or practice effects. This technique was recommended by several researchers (Butcher et al., 2006; Sireci & Berberoglu, 2000). Finally, the participants were asked about the authenticity of the translated adaptation of the PAI in a post-experimental questionnaire in an effort to compare cultural (ir)relevance. To conclude the study, participants were debriefed on the purpose of the study.

Analyses

The data of both the English and the French tests were compared using *t* values and Cohen's *d*, testing for differences between the two language versions of the test. We then analyzed reliability estimates of each scale in the English and French versions, by way of internal consistency and item inter-correlations. Further, the scale structure within each language group in the current sample was compared to the scale structure from the original U.S., Spanish, German and Greek samples mentioned earlier. Cronbach's α coefficients and mean inter-item correlations were calculated to compare scale structures.

Results

Mean differences between the English and French versions

Paired sample *t* tests were used to compare the data of both language versions (see Table 3), revealing significant differences after Bonferroni adjustment for multiple testing between the English and French versions on several scales and subscales. Relative to the English version, the French version showed a significantly higher *t* score on the PIM validity scale, with a moderate effect size ($d = 0.30$). In contrast, the English version showed a significantly higher score on the Borderline Features–Negative Relationships (BOR-N)

clinical subscale, displaying a medium effect size ($d = 0.37$).

Reliability estimates

To estimate the reliability of the different scales, we calculated the Cronbach's alpha and mean item inter-correlations (MICs) for the French and English versions. The internal consistency and MICs can be seen in Table 4. The validity scales displayed a similar pattern across language versions, though the English version exhibited higher reliability estimates (.29 to .69) than the French version (.29 to .56). None exceeded the 0.70 criterion for adequate Cronbach's alpha set by Nunnally (1978). The clinical scales exhibited similar patterns across language versions, with the English version displaying reliability estimates between −1.10 and .72, and the French version displaying reliability estimates between −.57 and .80. However, some of the scales produced low Cronbach's alpha in both versions. For example, a negative Cronbach's alpha was reported for the French Drug Problems (DRG) scale, and both the English and French Alcohol Problems (ALC) scales. The reliability estimates for the treatment and interpersonal scales ranged from .23 to .62 in the English version, and from .21 and .61 in the French version. None exceeded the 0.70 criterion.

MICs were below .20 for all scales on the French version, and were below .20 for all scales, except Negative Impression (NIM), Positive Impression (PIM) and Stress (STR) on the English version. Thus, the majority of the MICs are outside of the optimal range of .20 to .40 suggested by Briggs and Cheek (1986).

Since the publication of the PAI, several studies have examined the internal consistency of the PAI scales in diverse populations, including a bilingual Spanish sample, a German sample and a Greek sample. Table 5 summarizes the alpha coefficients reported in these studies. The results from the current study indicate that internal consistency

Table 3. Descriptive statistics for the scales and subscales of the PAI and differences between language versions.

Scale	Subscale	English M	English SD	French M	French SD	t	p[a]	Cohen's d
Validity								
INC		50.23	9.55	47.92	8.91	1.35	.184	0.25
INF		52.10	9.66	56.65	10.09	−3.23	.002	−0.46
NIM		49.80	9.16	51.02	8.40	−1.42	.161	−0.14
PIM		47.82	12.13	51.33	11.10	−3.84	.000*	−0.30
Clinical								
SOM		48.88	9.11	47.96	8.37	1.56	.125	0.11
	SOM-C	48.66	8.77	48.53	7.43	0.24	.815	0.02
	SOM-S	48.71	10.07	47.37	10.57	1.70	.098	0.13
	SOM-H	50.16	9.90	49.47	9.10	0.90	.372	0.07
ANX		53.90	11.67	53.88	9.37	0.03	.978	0.00
	ANX-C	55.21	11.64	54.55	8.50	0.58	.567	0.06
	ANX-A	52.59	13.07	51.26	11.18	1.05	.302	0.11
	ANX-P	53.82	10.39	54.92	9.39	−1.15	.259	−0.11
ARD		53.35	11.77	52.63	13.18	0.56	.576	0.06
	ARD-O	53.34	11.51	53.53	10.73	−0.20	.845	−0.02
	ARD-P	52.05	10.70	52.84	10.56	−1.01	.317	−0.07
	ARD-T	53.45	13.24	53.03	13.99	0.60	.552	0.03
DEP		51.92	10.64	51.69	10.08	0.33	.742	0.02
	DEP-C	55.66	12.54	54.97	10.94	0.66	.511	0.06
	DEP-A	53.03	12.11	51.53	10.96	1.68	.101	0.13
	DEP-P	49.21	9.07	48.87	9.74	0.38	.708	0.04
MAN		52.80	11.63	51.98	11.83	1.13	.265	0.07
	MAN-A	51.50	10.36	51.29	10.88	0.25	.801	0.02
	MAN-G	52.05	11.70	52.26	12.20	−0.21	.832	−0.02
	MAN-I	52.82	12.00	51.95	13.67	0.69	.498	0.07
PAR		53.76	9.71	54.57	9.30	−0.99	.327	−0.09
	PAR-H	59.39	11.82	57.95	11.31	1.21	.235	0.12
	PAR-P	50.79	11.04	51.11	9.61	−0.28	.778	−0.03
	PAR-R	0.08	11.01	51.97	11.32	−1.54	.133	−4.65
SCZ		51.90	9.85	51.22	8.84	0.81	.424	0.07
	SCZ-P	49.92	9.13	49.39	8.59	0.50	.618	0.06
	SCZ-S	51.55	11.12	52.24	9.55	−0.83	.411	−0.07
	SCZ-T	52.92	9.33	49.82	11.51	1.75	.089	0.30
BOR		55.14	11.82	53.08	11.54	2.79	.008	0.18
	BOR-A	51.39	11.78	51.26	12.41	0.12	.902	0.01
	BOR-I	58.92	12.59	55.08	14.33	2.23	.032	0.28
	BOR-N	54.63	11.71	50.63	9.77	4.04	.000*	0.37
	BOR-S	53.37	13.10	51.66	14.09	0.87	.393	0.13
ANT		54.14	11.72	55.41	12.03	−1.51	.137	−0.11
	ANT-A	50.39	10.77	49.71	10.66	0.67	.574	0.06
	ANT-E	53.79	10.47	57.11	14.75	−1.57	.124	−0.26
	ANT-S	55.00	13.33	58.58	13.24	−3.49	.001	−0.27
ALC		44.92	3.86	45.92	4.94	−1.64	.107	−0.23

(*Continued*)

Table 3. (*Continued*).

Scale	Subscale	English		French				Cohen's
		M	*SD*	*M*	*SD*	*t*	*p*[a]	*d*
DRG		48.73	8.93	47.35	7.35	1.54	.130	0.17
Treatment								
AGG		48.35	11.25	47.86	10.97	0.67	.504	0.04
	AGG-A	45.68	9.78	45.26	9.93	0.46	.646	0.04
	AGG-V	48.03	11.23	49.11	11.44	−1.40	.169	−0.10
	AGG-P	50.82	12.67	50.82	11.53	0.00	1.00	0.00
SUI		50.78	8.07	50.31	7.87	0.55	.585	0.06
STR		50.41	8.51	51.10	8.80	−0.81	.425	−0.08
NON		50.16	9.50	52.96	10.40	−3.57	.001	−0.28
RXR		51.80	8.93	51.96	8.64	−0.20	.843	−0.02
Interpersonal								
DOM		50.88	12.11	49.82	11.26	1.46	.151	0.09
WRM		49.76	10.06	47.59	9.52	2.67	.010	0.22

Note: PAI = Personality Assessment Inventory.
[a]Bonferroni adjustment for multiple testing reduces the significance level to .00094.
*The mean difference is significant after Bonferroni correction.

estimates resemble those from the previous studies for the validity scales. The average αs range from .46 to .98. However, the internal consistency estimates for the clinical, treatment and interpersonal scales tend to be lowest for the Canadian samples, with mean αs ranging from .48 to .92 in the clinical scales, .39 to .95 in the treatment scales, and .55 to .93 in the interpersonal scales.

Discussion

Since its introduction, the PAI has become one of the most widely employed measures of personality and psychopathology in the context of medical–legal examinations as it pertains to psychological injury (Boccaccini & Brodsky, 1999) and has been deemed one of the most widely accepted measures for a variety of forensic and psycho-legal applications (Lally, 2003). In this study, we examined the French-Canadian language adaptation of the PAI by assessing its psychometric properties in a sample of bilingual university students at a Canadian university. The mean PAI scale and subscale scores were compared between the two language versions, and the reliability estimates of each version were calculated and compared to diverse samples in the literature.

The comparison of responses to English and French-Canadian items presents several interpretational problems. The mean PAI scale and subscale scores revealed significant differences between the two test versions even after Bonferroni adjustment. The higher scores on the PIM validity scale in the French-Canadian version suggests that individuals taking the French-Canadian adaptation of the PAI are more likely to present with enhanced positive impression, trending towards invalid profiles (see Table 3). Conversely, these individuals are less likely to have attributes suggestive of borderline personality functioning. Clinicians using the French-Canadian adaptation of the PAI should be especially cautious, in that these disparities in scale and subscale scores arise when controlling for individual and cultural differences (i.e. the differences are likely due to the linguistic aspects of the items). It must be noted that these comparisons were exploratory in nature. As such, the high number of mean comparisons resulted in an extremely

Table 4. Cronbach's alpha (α) and mean item inter-correlations for the PAI scales of the French and English samples.

Scale	N (Items)	French		English	
		α	MIC	α	MIC
Validity					
INC	20	.56	.05	.45	.03
INF	8	.29	.03	.29	.03
NIM	9	.54	.11	.68	.23
PIM	9	.53	.11	.69	.20
Clinical					
SOM	24	.64	.11	.71	.14
ANX	24	.49	.06	.61	.08
ARD	24	.74	.12	.70	.10
DEP	24	.55	.13	.58	.08
MAN	24	.80	.16	.83	.17
PAR	24	.67	.08	.51	.11
SCZ	24	.60	.08	.68	.09
BOR	24	.70	.09	.72	.09
ANT	24	.74	.12	.72	.13
ALC	12	-.57	.05	−1.10	-.06
DRG	12	-.06	.14	.47	.18
Treatment					
AGG	18	.47	.08	.56	.09
SUI	12	.34	.10	.47	.14
STR	8	.56	.15	.64	.20
NON	8	.38	.07	.23	.04
RXR	8	.21	.02	.40	.07
Interpersonal					
DOM	12	.61	.10	.47	.06
WRM	12	.52	.08	.62	.12
M		.47	.09	.50	.11

Note: PAI = Personality Assessment Inventory; MIC = mean item inter-correlation.

conservative Bonferroni adjustment, and a high rate of false negatives is probable. Future studies should focus on targeting specific scales.

The internal consistency of the French-Canadian and English PAIs was also examined. The PAI scales demonstrated poor internal reliability in both language versions, as evidenced by Cronbach alpha coefficients below the criterion standards suggested by Nunnally (1978). Morey (2007) suggests that low internal consistency estimates are expected of INC and INF validity scales, as they were not intended to measure substantive theoretical constructs, but rather are composed of error variance. However, the remaining scales are theoretically driven and therefore should have adequate internal consistency estimates. Yet, none of the validity, treatment or interpersonal scales were $\geq .70$, and just four clinical scales on the French-Canadian (i.e. Anxiety-Related Disorders, ARD; Mania, MAN; Borderline Features, BOR; Antisocial Features, ANT) and five clinical scales on the English (i.e. Somatic Complaints, SOM; ARD; MAN; BOR; ANT) versions of the PAI were above this standard. In addition, the negative internal consistency estimates

Table 5. Internal consistency (α) coefficients for PAI scales: comparing past studies with the current study.

| | United States[a] Original English | | | United States[b] | | Germany[c] German | | Greece[d] Greek | | Canada | |
	Normative (N = 1000)	University students (N = 1050)	Clinical patients (N = 1246)	English Outpatients (N = 69)	Spanish Outpatients (N = 69)	Normative N = 749	Both inpatients and outpatients (N = 450)	Outpatients (N = 300)	Non-clinical (N = 1120)	French-Canadian University students (N = 50)	Original English University students (N = 50)
Validity											
INC	.45	.26	.23	.22	.44	.71	–	–	–	.56	.45
INF	.52	.22	.40	.55	.29	.26	–	–	–	.29	.29
NIM	.72	.63	.74	.66	.54	.68	.86	.97	.85	.54	.68
PIM	.71	.73	.77	.75	.70	.71	.94	.99	.86	.53	.69
M	.60	.46	.54	.55	.49	.59	.90	.98	.86	.48	.53
Clinical											
SOM	.89	.83	.92	.81	.71	.91	.90	.90	.79	.64	.71
ANX	.90	.89	.94	.79	.80	.89	.78	.98	.82	.49	.61
ARD	.76	.80	.86	.71	.69	.74	.80	.98	.90	.74	.70
DEP	.87	.87	.93	.74	.72	.88	.80	.96	.80	.55	.58
MAN	.82	.82	.82	.87	.66	.81	.84	.86	.95	.80	.83
PAR	.85	.86	.89	.78	.74	.83	.82	.96	.84	.67	.51
SCZ	.81	.82	.89	.73	.66	.80	.89	.97	.82	.60	.68
BOR	.87	.86	.91	.80	.82	.84	.89	.99	.74	.70	.72
ANT	.84	.86	.86	.79	.57	.82	.86	.76	.85	.74	.72
ALC	.84	.83	.93	.51	.66	.82	.86	.85	.87	–.57	–1.10
DRG	.74	.66	.89	.86	.40	.63	.96	.87	.80	–.06	.47
M	.84	.83	.89	.76	.68	.82	.85	.92	.83	.48	.49
Treatment											
AGG	.85	.89	.90	.69	.82	.81	.97	.80	.83	.47	.56
SUI	.85	.87	.93	.81	.46	.87	.86	.84	.92	.34	.47
STR	.76	.69	.79	.72	.40	.74	.96	.82	.74	.56	.64
NON	.72	.75	.80	.64	.71	.72	.97	.89	.86	.38	.23
RXR	.76	.72	.80	.25	.75	.70	.99	.89	.84	.21	.40
M	.79	.78	.84	.62	.63	.77	.95	.85	.84	.39	.46
Interpersonal											
DOM	.78	.81	.82	.54	.41	.72	.76	.86	.83	.61	.47
WRM	.79	.80	.83	.68	.71	.76	.85	.99	.83	.52	.62
M	.79	.81	.83	.61	.56	.74	.81	.93	.83	.57	.55

Note: PAI = Personality Assessment Inventory.
[a]Morey, 2007.
[b]Rogers et al., 1995.
[c]Groves & Engel, 2007.
[d]Lyrakos, 2011.

demonstrate that the correlations between the items are weak among the ALC and DRG scales, and thus should be considered separate entities entirely. Another more likely explanation for these negative internal consistency scores is that although the covariances among items may be positive, sampling error has produced a negative average in a given number of cases due to the small sample size (i.e. there is greater within-subject variability than between-subject variability). Overall, these results are discrepant with the normative U.S. sample (Morey, 2007), where most clinical, treatment and interpersonal scales exceeded Nunnally's criterion standards. A plausible explanation is that the current sample consisted of bilingual individuals, which is typically a select sample and may not be generalizable to monolingual samples due to factors such as acculturation and education (Fernandez, Boccaccini, & Noland, 2008; Sireci & Berberoglu, 2000). To summarize, the internal consistency coefficients demonstrated in this sample were unacceptably low, suggesting a high proportion of random or otherwise errorful responding. These results suggest that the scores on the scales and subscales of the PAI be interpreted with caution pending further research on the topic.

A second determinant of reliability is item content homogeneity and whether the items in a scale cover many different aspects, or focus on only a few. To estimate this, MICs were reported for each of the scales. MICs differ from the internal consistency estimates, in that they are not influenced by scale length and are therefore a clearer measure of homogeneity. Briggs and Cheek (1986) state that the optimal level of homogeneity occurs when the MIC is in the .20 to .40 range. Since the MICs were below .20 for all scales on the French version, and were below .20 for all scales except NIM, PIM and STR on the English version, the items are likely too broad to adequately represent a single total scale score. Though the PAI was meant to measure relatively broad classes of behavior, many of the MICs were below the

lower threshold of .10, demonstrating poor item homogeneity.

Limitations of the study are related to the bilingual sample and lack of time interval between test version administration. Bilinguals are rarely homogeneous with respect to their language proficiencies, resulting in a differential response to the same item due to a misunderstanding of the item in the less familiar language. Though the use of bilingual test-takers allows items to be flagged across languages, Sireci and Berberoglu (2000) stated that 'bilingual test takers cannot be used to unequivocally determine whether items function equivalently across languages' (p. 244). To make meaningful cross-cultural comparisons, more extensive studies are needed with a larger sample size and using differential item functioning to assess measurement invariance. Moreover, future studies should assess the reliability of the French-Canadian translation in a French-Canadian university in the attempt to discern the cultural and clinical heterogeneity among French-Canadian populations. In addition, a large time interval to remove the variance stemming from immediate administration was not employed in this study. However, it can be argued that allowing for a larger time interval between the administrations could potentially add a state-effect. In other words, the participant's psychological state could vary between administrations, potentially confounding the results. In this study, the potential effects of psychological state were controlled for due to the immediate re-administration of the PAI (either translated or English version). Given our findings that differences in French-Canadian and English versions exist for the PAI, a more extensive study will be undertaken using the gold standard test–retest time interval suggested in the literature (Chen & Bond, 2010; Sireci & Berberoglu, 2000).

Overall, the results of this study suggest that we must be cautious when using the French-Canadian adaptation of the PAI. The lack of reliability in the French translation and the lack of item homogeneity require explanation. Scale reliability is commonly said

to limit validity and is critical for test interpretation. As such, high internal consistency estimates and adequate MICs are essential when scores are used for making important decisions about individuals, such as those made in medicolegal examinations. While self-report measures are quick, inexpensive and easy to administer and score, clinicians and researchers must be sure that patients are able to understand and correctly interpret the intention of the items before test scores can be considered an accurate representation of the individual's functioning. Based on the findings of the current study, the clinical usefulness of the French-Canadian PAI appears to be premature due to poor reliability. The utility of the French-Canadian PAI is questionable at this time and should be interpreted only with great caution, particularly in a medical–legal context.

Ethical standards

Declaration of conflicts of interest

Eliyas Jeffay has declared no conflicts of interest

Angela Sekely has declared no conflicts of interest

Michel Lacerte has declared no conflicts of interest

Konstantine K. Zakzanis has declared no conflicts of interest

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Boccaccini, M.T., & Brodsky, S.L. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice*, *30*(3), 253–259. doi:10.1037/0735-7028.30.3.253

Briggs, S.R., & Cheek, J.M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*(1), 106–148. doi:10.1111/j.1467-6494.1986.tb00391.x

Butcher, J., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Butcher, J., Derksen, J., Sloore, H., & Sirigatti, S. (2003). Objective personality assessment of people in diverse cultures: European adaptations of the MMPI-2. *Behaviour Research and Therapy*, *41*(7), 819–840. doi:10.1016/S0005-7967(02)00186-9

Butcher, J.N., Mosch, S.C., Tsai, J., & Nezami, E. (2006). *Cross-cultural applications of the MMPI-2*. Washington, DC: American Psychological Association.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada (2010). *Tri-council policy statement: Ethical conduct for research involving humans.* Ottawa, ON: Interagency Secretariat on Research Ethics.

Candell, G., & Hulin, C. (1986). Cross-language and cross-cultural comparisons in scale translations. *Journal of Cross-Cultural Psychology*, *17*(4), 417–440. doi:10.1177/0022002186017004003

Chen, S.X., & Bond, M.H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528. doi:10.1177/0146167210385360

Cheung, F.M. (2009). The cultural perspective in personality. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (p. 44). New York, NY: Oxford University Press.

Cheung, F.M., Leung, K., Fan, R.M., Song, W., Zhang, J., & Zhang, J. (1996). Development of the Chinese personality assessment inventory. *Journal of Cross-Cultural Psychology*, *27*(2), 181–199. doi:10.1177/0022022196272003

Ellis, B.B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology*, *74*(6), 912–921. doi:10.1037/0021-9010.74.6.912

Fernandez, K., Boccaccini, M. T., & Noland, R. M. (2008). Detecting over- and underreporting of psychopathology with the Spanish-language personality assessment inventory: Findings from a simulation study with bilingual speakers. *Psychological Assessment*, *20*(2), 189–194. doi:10.1037/1040-3590.20.2.189

Geisinger, K.F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*(4), 304–312. doi:10.1037/1040-3590.6.4.304

Groves, J.A., & Engel, R.R. (2007). The German adaptation and standardization of the personality assessment inventory (PAI). *Journal of Personality Assessment*, *88*(1), 49–56. doi:10.1080/00223890709336834

Jeanrie, C., & Bertrand, R. (1999). Translating tests with the international test commission's guidelines: Keeping validity in mind. *European Journal of Psychological Assessment*, *15*(3), 277–283. doi:10.1027//1015-5759.15.3.277

Lally, S.J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, *34*(5), 491–498. doi:10.1037/0735-7028.34.5.491

Lyrakos, D.G. (2011). The development of the Greek personality assessment inventory. *Psychology*, *02* (08), 797–803. doi:10.4236/psych.2011.28122

Morey, L. (1991). *Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. (2007). *Personality Assessment Inventory (PAI): Professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

PAI Software Portfolio (2000). *PAI Software Portfolio Windows. (Version 2)*. Odessa, FL: Psychological Assessment Resources Inc. https://www.parinc.com/products/pkey/295

Rogers, R., Flores, J., Ustad, K., & Sewell, K.W. (1995). Initial validation of the personality assessment Inventory—Spanish version with clients from Mexican American communities. *Journal of Personality Assessment*, *64*(2), 340–348. doi:10.1207/s15327752jpa6402_12

Sireci, S.G., & Berberoglu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, *13*(3), 229–248. doi:10.1207/S15324818AME1303_1

Stover, J.B., Solano, A.C., & Liporace, M.F. (2015). Personality Assessment Inventory: Psychometric analyses of its Argentinean version. *Psychological Reports*, *117*(3), 799–823. doi:10.2466/08.03.PR0.117c27z2

Wilkinson, G.S., & Robertson, G.J. (2006). *Wide range achievement test–fourth edition*. Lutz, FL: Psychological Assessment Resources.